

DOCUMENT RESUME

ED 406 438

TM 026 417

AUTHOR Whitaker, Jean S.
TITLE Interpretation of Structure Coefficients Can Prevent Erroneous Conclusions about Regression Results.
PUB DATE 25 Jan 97
NOTE 18p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, January 23-25, 1997).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Correlation; Factor Analysis; Predictor Variables; *Regression (Statistics); *Research Methodology; Research Problems
IDENTIFIERS Beta Weights; *Structure Coefficients

ABSTRACT

The increased use of multiple regression analysis in research warrants closer examination of the coefficients produced in these analyses, especially ones which are often ignored, such as structure coefficients. Structure coefficients are bivariate correlation coefficients between a predictor variable and the synthetic variable. When predictor variables are correlated with each other, regression results may be seriously distorted by failure to interpret structure coefficients. Structure coefficients have analogues in all analyses (e.g., canonical analysis, factor analysis, and discriminant analysis) and should be interpreted in all analyses. Several examples of research in which not examining structure coefficients led to misinterpreting results are cited. A small heuristic example is presented to provide a concrete example of how interpretation of regression results might differ when predictor variables are correlated with each other. The astute researcher should examine both beta weights and structure coefficients when interpreting regression results with correlated predictor variables. (Contains 1 figure, 1 table, and 27 references.) (Author/SLD)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

Running head: INTERPRETATION OF STRUCTURE COEFFICIENTS

ED 406 438

Interpretation of Structure Coefficients Can Prevent Erroneous

Conclusions About Regression Results

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to
improve reproduction quality.

• Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY

JEAN S. WHITAKER

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

Jean S. Whitaker

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research

Association, Austin, TX, January 25, 1997.

Tm026417

Abstract

The increased use of multiple regression analysis in research warrants closer examination of the coefficients produced in those analyses, especially ones which are often ignored, such as structure coefficients. Structure coefficients are bivariate correlation coefficients between a predictor variable and the synthetic variable, YHAT. When predictor variables are correlated with each other, regression results may be seriously distorted by failure to interpret structure coefficients. Structure coefficients have analogs in all analyses (e.g., canonical analysis, factor analysis and discriminant analysis) and should be interpreted in all analyses. Several examples of research in which structure coefficients were not examined and subsequently misinterpreted results are cited. A small heuristic example is presented to provide a concrete example of how interpretation of regression results might differ when predictor variables are correlated with each other. The astute researcher should examine both beta weights and structure coefficients when interpreting regression results with correlated predictor variables.

Interpretation of Structure Coefficients Can Prevent Erroneous

Conclusions About Regression Results

Willson (1980) examined the research literature over a ten year span (1969-1978) and documented the increased use of multiple regression analysis in research designs. He also noted the inclusion of multiple regression analysis in more than half of the education research texts published during that same time period. The utility of multiple regression analysis in behavioral science applications is well established (Thompson, 1985). The versatility of multiple regression analysis is most evident in the "amount of information it yields about relationships among variables" (Gall, Borg, & Gall, 1996, p. 434). The authors also noted that "it can be used to analyze data from any of the major quantitative research designs.... And it provides estimates both of the magnitude and statistical significance of relationships between variables" (pp. 434).

Kerlinger and Pedhazur (1973, p. 3) pointed out that multiple regression analysis "can be used equally well in experimental or nonexperimental research. It can handle continuous and categorical variables. It can handle two, three, four, or more independent variable. ...It can do anything the analysis of variances does - sum of squares, mean squares, *F* ratios - and more."

Although multiple regression enjoys superiority over some other methods, as a general analytic procedure, it is related to other parametric methods by the concept of a general linear model (Baggaley, 1981; Cohen, 1968). Several researchers (Baggaley, 1981; Cohen, 1983; Knapp, 1978) have recognized that all parametric methods such as *t*-

tests, ANOVA, ANCOVA, MANOVA, MANCOVA, and discriminant analysis are actually special cases of canonical correlation analysis and, therefore, interrelated.

Some of the evidence for the utility of interpreting structure coefficients in factor analysis (Gorsuch, 1983; Weigle & Snow, 1995), regression (Bowling, 1993; Daniel, 1990; Perry, 1990; Thompson & Borriello, 1985), discriminant analysis (Pedhazur, 1982) and canonical analysis (Meredith, 1964; Thompson, 1984, 1988) has suggested that doing so may be essential in univariate analyses (Friedrich, 1991). However, Friedrich (p. 1) also noted that structure coefficients "are not routinely reported and utilized in the interpretation of such [parametric] analyses."

Some researchers regard structure coefficient interpretation as useless. Harris (1992) maintained that interpretation of structure coefficients can be misleading and, therefore, should be rejected in favor of interpretation of scoring coefficients, especially in multiple regression and its special case, i.e., two-group discriminant analysis. Pedhazur (1982, p. 691) discussed the utility of structure coefficient interpretation for factor analysis, discriminant analysis and canonical analysis, and he indicated that structure coefficients "do not enhance the interpretation of results of multiple regression analysis." He also added that, "such coefficients [structure coefficients] are simply zero-order correlations of independent variables with the dependent variable divided by a constant, namely, the multiple correlation coefficient. Hence, the zero-order correlations provide the same information." Thompson and Borriello (1985, p. 208) responded to Pedhazur by suggesting that, "...interpretation of only the bivariate correlations seems counterintuitive. It appears inconsistent to first declare interest in an omnibus system of variables and then to consult values that consider the variable taken only two at a time."

The present paper introduces the reader to the concept of multiple regression and the importance of interpreting structure coefficients when correlations between predictor variables exist. A brief overview of multiple regression analysis is given first, followed by a discussion of collinearity and structure coefficients. Examples from the research literature that include interpretation of structure coefficients are cited, and a small heuristic data set is presented to illustrate concretely how interpretation of regression results might differ when predictor variables are correlated with each other.

Multiple Regression

Multiple regression is an extension of the concept of simple regression, which relates directly to correlation analysis (Kachigan, 1986). According to Kerlinger and Pedhazur (1973), the r used to indicate the coefficient of correlation really means regression. It is said that we study the regression of Y scores on X scores. Specifically, we are trying to predict Y from X , but we could just as easily be trying to predict X from Y , i.e., in the bivariate case, the assignment of X and Y is arbitrary. Predictive ability is increased as the correlation between two variables increases. The concepts of correlation and a straight line can be used to develop the concept of linear regression (Hinkle, Wiersma, & Jurs, 1994). The regression equation for a sample is:

$$Y_i = a + bX + e$$

where Y_i = score of individual i ; a = the value of X on the Y intercept; b = regression coefficient (weight) or the slope of the regression line; e = error for individual i (difference between each person's actual score (Y) and their predicted score (Y^*)). If " e " is equal to zero, then, every Y equals Y^* for each individual. In the above equation the two conventional regression weights are applied, i.e., a is the additive constant and applied

to every case and "b" is the multiplicative constant which is applied to the predictor variable for each case (Thompson, 1992).

In multiple regression analysis the above formula is extended to include more than one predictor variable. The concept of least squares is operational here also and is used to develop an equation wherein predictor variables (e.g., $X_1, X_2, \dots, X_n$) are optimally weighted so as to minimize the distance (i.e., the e or residual score) between each individual's predicted score, $\hat{Y}$, and their actual score, Y . The regression equation generally takes the following form:

$$Y = \hat{Y} = a + b_1(X_1) + b_2(X_2)$$

The "a" weight equals the $\hat{Y}$ score when X_1 and X_2 are both equal to zero. Regression coefficients, b weights, for the independent variables X_1 and X_2 are designated b_1 and b_2 , respectively. Actually, a , b_1 , and b_2 , are all employed in the least squares method. The "b" weights are sensitive to the correlation of each predictor variable with Y , the correlation among predictor variables, and the variability of predictor variables in relation to the dependent variable, Y . These sensitivities create problems in interpreting b weights.

Regression coefficients are usually standardized in order to facilitate comparison across variables with different standard deviations, scales, or metrics. Typically, b weights are standardized prior to interpretation of regression results. The "b" weights can be converted to standardized weights, called β weights, using the following formula:

$$\beta = \frac{SD_x}{SD_y} \text{ "b"}$$

The "b" and β weights will be equal when either is zero or the standard deviations of both variables are equal (Thompson, 1992). As Thompson (1994) explained, "The β weights in a regression analysis are the correlation coefficients between the respective predictors and the dependent variable only when those predictors that are correlated with the dependent variable are perfectly uncorrelated with each other" (p. 30).

Typically, researchers judge the relative contribution of each of the predictor variables in the regression equation based on the magnitude of their beta weights (Cooley & Lohnes, 1971). The unwary researcher might be tempted to regard the predictor variable with the largest absolute value as the greatest predictor. As Figure 1 demonstrates, it is possible to have a predictor variable with the greatest predictive potential lose credit to two (or more) other predictors whose predictive area overlaps that of the first predictor. The first predictor is given no credit for predictive potential and could have a beta weight of zero. In this instance, it is important to have information about the true predictive potential of that variable, information that can be easily gained by examining each predictor variable's structure coefficient.

Collinearity

When predictor variables are correlated with each other the terms "collinearity," "multicollinearity," or "ill conditioning" may be used as descriptors (Thompson & Borrello, 1985). Numerous researchers have cautioned their readers about the complexity collinearity introduces into both least squares calculations (Belsley, Kuh, & Welsh, 1980), statistical accuracy of test statistics (Pedhazur, 1982), and interpretations of multiple regression results (Pedhazur, 1982). Some researchers have suggested that

collinearity be avoided in the original design choices. However, as Thompson and Borrello (1985, p. 204) noted, in certain cases "collinearity reflects sound design decisions of the researcher. Researchers purposely introduce collinearity when using multiple measures of variables in which they have greater interest or which are more important from a theoretical point of view."

Some have suggested that collinearity may more realistically reflect the underlying nature of the constructs under study (Belsley et al., 1980). In reference to canonical correlation, Meredith (1964, p. 55) stated that, "If variables within each set are moderately intercorrelated the possibility of interpreting the canonical variates by inspection of the appropriate regression weights is practically nil." It is apparent that collinearity can create problems in multivariate analysis, however, collinearity may not be problematic when it reflects the reality of the researchers inquest. Logically, it would seem to be in the researcher's best interest to have a clear idea of their research question(s) and to attempt to understand how their results answered those questions. As Thompson (1992, p. 16) noted, "...the utility of statistics varies somewhat from problem to problem or situation to situation."

Despite the information provided by interpreting the structure coefficients, researchers do not all agree on this point. Harris (1992) has argued vehemently against the interpretation of emergent variables on the basis of structure coefficients, especially for multiple regression. On the other side of the debate researchers like Thompson (1992, p. 14) recommend that "... the thoughtful researcher should always interpret either (a) both the beta weights and the structure coefficients or (b) both the beta weights and the bivariate correlations of the predictors with Y." Thompson and Borrello (1985)

suggested early in the debate that beta weights, structure coefficients, and zero-order correlations are important aids to interpretation. Pedhazur (1982, p. 691) argued that structure coefficients "are simply zero-order correlations of independent variables with the dependent variable divided by a constant, namely, the multiple correlation coefficient. Hence, the zero-order correlations provide the same information." However, as Thompson and Borrello (1985, p. 208) stated that "it must be noted that interpretation of only the bivariate correlations seems counterintuitive. It appears inconsistent to first declare interest in an omnibus system of variables and then to consult values that consider the variables taken only two at a time."

Structure Coefficients

Structure coefficients are not affected by collinearity. A structure coefficient is the correlation between an independent variable and the vector of composite scores obtained by applying the regression equation to subjects' scores on the independent variables (Pedhazur, 1982).

When predictor variables are perfectly uncorrelated, the structure coefficient yields the same interpretation as the beta weight or the individual correlation of predictor variable with Y^* . Also, in this hybrid case "the sum of the r^2 's for the predictors (each representing how much of the dependent variable a predictor can explain) will equal the R^2 involving all the predictors..." (Thompson, 1992, p. 12).

Thorndike (1978) indicated that structure coefficients honor the reality of the relationship of variables under study. As Thompson (1994) explained, "In regression analyses, to avoid result misinterpretation, both standardized weights and structure coefficients, or, both standardized weights and correlation coefficients between the

predictor variables and the dependent variable, should always be presented together" (p. 20).

Examples from the research literature

Several examples from the research literature are offered as concrete examples of instances where interpretation of structure coefficients enhances or lends clarity to the reality of the data. Daniel (1990) determined that the use of structure coefficients was superior to other methods in his analysis of MANOVA results because they honored the multivariate reality of the data, minimized experiment-wise Type I error rates, and were neither inflated or suppressed by collinearity among variables. Daniel's study included 36 subjects in a one-way design with experimental condition (3 levels of the predictor variable) and 3 continuous criterion variables (scores on 3 subtests in an achievement battery). When he consulted the two sets of function coefficients, SCORE3 weighted most heavily on the first function, SCORE1 on the second function, and SCORE2 had a near zero weight which lead him to conclude that it contributed to neither function. However, examination of the structure coefficients showed SCORE1 and SCORE2 weighting on the second synthetic variable. Thus, both analyses identified two distinct constructs, but the second construct was interpreted differently when structure coefficients were used.

In 1990, Perry presented regression results from the "Heart Smart" study. Her dependent variable was HDL cholesterol ("good" cholesterol) and her predictor variables were are follows: MILESEC, time required to walk/run one mile; SYSTOLAV, average of 6 systolic pressure readings; and, POND, ponderosity (ratio of weight in kilograms to cubed value of height in centimeters. Perry noted that beta weight interpretation would

suggest that POND had little or no predictive value, whereas examination of the structure coefficients indicated that POND had virtually the same predictive power as MILESEC.

Bowling (1993) examined over 20 research based articles using multiple regression that were published in the Journal of Counseling Psychology between January, 1990 and April, 1993. Only three reported structure coefficients. Bowling concluded that, "decisions related to program funding, interventions, and general understanding of human behavior may all be misdirected when structure coefficients are not computed as part of regression analyses, unless the predictor variables are perfectly uncorrelated" (p. 12).

Heuristic Data

A small heuristic example is presented in Table 1 to illustrate the importance of interpreting structure coefficients in regression results. The example is a multiple regression analysis with three independent variables (i.e., X_1 , X_2 , and X_3). Examination of beta weights only would suggest that X_1 contributed the most to the regression equation and represented the best predictor. X_3 's seemingly insignificant beta weight would suggest that it contributed little to the regression equation. However, examination of each of the three predictors' structure coefficients provides a different interpretation of the actual predictive potential of the three variables. There is also evidence of multicollinearity. As Thompson (1992) stated "When all predictors have nonzero betas and nonzero structure coefficients (or r 's with Y), then predictor variables overlap with each other, i.e., are multicollinear" (p. 17).

Summary

The serious researcher has a professional and ethical responsibility to disseminate complete information and accurately interpret findings to fellow colleagues and

consumers of research, especially those in decision-making positions. In light of the increasing use of multiple regression in educational research and the potential for different interpretations, it would seem that structure coefficients should be examined whenever collinearity between predictor variables exists. The most basic purpose of statistics is to understand our data. How blatant is our error when we ignore information that is crucial to our understanding, conclusions, and decision-making?

References

Belsley, D.A., Kuh, E., & Welsh, R. E. (1980). Regression diagnostics: Identifying influential data and sources of collinearity. New York: John Wiley & Sons.

Baggaley, A. R. (1981). Multivariate analysis: An introduction for consumers of behavioral research. Evaluation Review, 5, 123-131.

Bowling, J. (1993, November). The importance of structure coefficients as against beta weights: Comments with examples from the counseling psychology literature. Paper presented at the annual meeting of Mid-South Education Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 364 606)

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443. (48, 378-399)

Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences.

Cooley, W. W., & Lohnes, P. R. (1971). Multivariate data analysis. New York: John Wiley & Sons.

Daniel, L. G. (1990, January). The use of structure coefficients in multivariate educational research: A heuristic example. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin. (ERIC Document Reproduction Service No. ED 315 451)

Friedrich, K. R. (1991, April). The importance of structure coefficients in parametric analysis. Paper presented at the annual meeting of the American Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED 330 725)

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). Educational research: an introduction (6th ed.). New York: Longman Publishers USA.

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Erlbaum.

Harris, R. J. (1992, April). Structure coefficients versus scoring coefficients as bases for interpreting emergent variables in multiple regression and related techniques. Paper presented at the annual meeting of the American educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 356 231)

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1994). Applied statistics for the behavioral sciences (3rd ed.). Boston: Houghton Mifflin Company.

Huck, S. W. & Cormier, W. H. (1996). Reading statistics and research (2nd ed.). New York: Harper and Row.

Kachigan, S. K. (1986). Statistical analysis. New York: Radius Press.

Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart, & Winston.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological Bulletin, 85, 410-416.

Meredith, W. (1966). Canonical correlations with fallible data. Psychometrika, 29, 55-65.

Pedhazur, E. J. (1982). Multiple regression in behavioral research (2nd ed.). New York: Holt, Rinehart, & Winston.

Perry, L. (1990, January). The use of structure coefficients in regression research. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin. (ERIC Document Reproduction Service No. ED 315 448)

Seibel, H. P., Wallbrown, F. H., Reuter, E. K., & Barnett, R. W. (1990). Further evidence concerning motivational distortion on the sixteen personality factor primaries by male felons. Journal of Personality Assessment, 55, 367-375.

Thompson, B., & Borrello, G. M. (1985). The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Thompson, B. (1980, April). Canonical correlation: Recent extensions for modeling educational processes.

Thompson, B. (1984). Canonical correlation analysis: Uses and interpretation. Newbury Park: SAGE.

Thompson, B. (1988, April). Canonical correlation analysis: An explanation with comments on correct practice. Paper presented at the annual meeting of the American Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 295 957)

Thompson, B. (1992, April). Interpreting regression results: beta weights and structure coefficients are both important. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 344 897)

Thompson, B. (1994, April). Common methodology mistakes in dissertations, revisited. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 368 771)

Willson, V. L. (1980). Research techniques in AERJ articles: 1969 to 1978. Educational Researcher, 9, 5-10.

Figure 1.

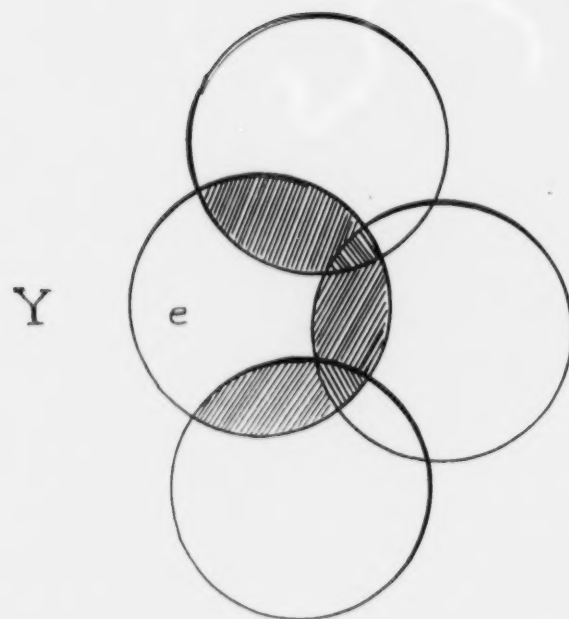


Table 1

Regression Results

<u>Variable</u>	<u>Regression Beta Weights</u>	<u>Structure Coefficients</u>
X_1	0.61748	0.63110
X_2	0.45671	0.34526
X_3	0.03424	0.62899
